[SQUEAKING]

[RUSTLING]

[CLICKING]

SARA ELLISON: OK. Let's go ahead and get started. Today, I'm going to talk about-- start talking about random variables and distributions of random variables. And at the end, I will probably also get to a discussion of joint distributions.

Tuesday, then, Esther is going to be lecturing. She's going to do a bunch of examples and also talk about, as she mentioned last time, some data sources and techniques for gathering data, such as web scraping. And then, she's going to play around with data little bit and do some techniques that essentially used histograms, kernel density plots, so forth-- some techniques that we essentially use to estimate or approximate things like distributions of random variables.

And then Wednesday, I'll be lecturing again. We'll talk about independence of random variables, conditional distributions, and maybe some other stuff, too.

The recitation on Friday is going to be continuation of the R tutorial. And there could be-- depending on exactly what gets put on the problem set, there could be a couple little sort of hints about issues in probability that you might need for the problem set in the recitation on Friday. But it's mostly going to be on R. Questions before I get started?

So as I said, I'm going to start talking about this sort of mathematical construct called random variables. So oftentimes, there's some numerical characteristic in the sample space that we're interested in. So we've talked about a lot of-- we've seen examples of different sample spaces.

And some of them have-- we can talk about a numerical characteristic of the sample space that we might be interested in, like the sum of the faces of two dice, or the number of 3-point field goals that Steph Curry makes in his next six attempts, the number of vegetarian toppings I get on my pizza. And there's an important and useful mathematical construct that we use to analyze those numerical characteristics, and it's called a random variable.

So a random variable is a real valued function whose domain is the sample space. And, as I said, it sort of-- it's real valued. So it goes from, in particular, the sample space to the real line.

So I don't know if this graphic sort of helps you think about the relationship between a probability and a random variable in a sample space, and the unit interval and the real line. It's all-- this is probably still pretty abstract to you, and we'll go through examples. But at least for me, this kind of graphic helps me think about what the difference between a probability, like we defined last time, is and a random variable.

So a probability goes from the set of all subsets of the sample space into the unit interval. So it always has values between 0 and 1, as we saw last time. A random variable is-- it goes from the sample space to the real line. And it's just some numerical characteristic of the sample space that we care about.

And the probability on the left induces the distribution of our random variable. So I'll be more clear about that when we go through examples. And that will become much more concrete. But basically, we just think about the probability that exists that's defined on the sample space as inducing a distribution of our random variable.

So all of the examples we've seen so far, all of the probability examples, would give rise to a type of random variable called discrete. And that's a random variable that takes on a finite or countably infinite number of values.

So the examples that I gave you on the previous slide-- the number of vegetarian toppings that I'm getting-- it's not going to be-- it's going to be an integer value, for instance. Discrete random variables don't have to all be integers. But something that can only take on the value of integers, for instance, would be a discrete random variable.

We can also, however, consider a natural generalization or, kind of depending on the mathematical framework you're using, a different flavor of this construct. And that's allowing a random variable to take on any value in some interval, bounded or unbounded, of the real line. And this is called a continuous random variable.

So, in this class, and in most-- in a lot of statistics classes, in fact, we will mostly deal with continuous random variables. And there are a couple reasons for this. This is partly because there are lots of discrete random variables in the real world that can be adequately approximated with continuous random variables. And so then that's sort of handy if we can do everything in the framework of a continuous random variable.

The other thing is that sometimes the math is just simply easier to do when you're considering complicated functions of random variables, for instance. Sometimes using continuous random variables is just a lot simpler than doing everything with discrete random variables.

So we will-- today, we're going to see examples of discrete random variables, continuous random variables, and we'll continue to see them throughout the semester. But as we go through the semester, we'll see more and more examples of continuous random variables.

So for a discrete random variable, we can often start with a verbal description. And then based on that verbal description-- a verbal description of some kind of an experiment. And then based on that verbal description, we can calculate probabilities for each value of the random variable. And then write down a function or draw a graph or something like that that describes the probabilities for different values of the random variable. And this is called a probability function.

Now, last time when we saw the hypergeometric example and the binomial example, we did the first part of this exercise already. We hadn't defined random variables yet, but we did the first part of this exercise. And so now the two examples I'm going to do are sort of completing this exercise for hypergeometric and binomial random variables.

So back to the pizza example. So we're going to define a random variable x. And x is the number of vegetarian toppings I get on my pizza if I draw the area for toppings randomly without replacement. So remember, a random variable is some numerical characteristic of the sample space we care about-- in this case, number of vegetarian toppings.

So last time we calculated the probability-- I can't remember exactly which probability we calculated-- the probability that I got one vegetarian topping when I drew two at random, when there were six nonvegetarian and five vegetarian or something like that. So we calculated that probability.

But what I want to do now is I want to calculate those probabilities for all of the various possible values of x. So instead of just saying, what's the probability that I get one vegetarian topping, I want to calculate the probability that I get zero, that I get one, that I get two, and so forth up to the maximum number of vegetarian topics I can get.

And, in this case, the maximum number is the max of 6 and n. Why is that? So perhaps you don't remember what the setup of this problem was. 6 is the number of vegetarian toppings available. So I can't get any more than that. n is the number of toppings that I'm choosing at random. So I can't get any more than that.

So the-- oh, sorry. Up to a-- sorry-- the minimum. Or yeah. Yeah, the minimum of 6 and n. Yeah. Sorry. I guess I wrote this in a somewhat confusing way, but I think it's actually correct up there. So up to 6 or n, whichever is smaller.

So I use this formula. And a couple things to note about it. It's a little bit different from the one I had up on the screen last time. First of all-- well, I changed the notation. First of all, I changed n1 to x. And that's just to be consistent with my notation for the random variable.

And then I noted that n1 plus n2 was equal to n. And so then I just-- since I substituted x for n1, then I just substituted n minus x for n2. So just changed the notation a little bit. Otherwise, it's exactly the same formula.

And then the other thing I want to note that I forgot to note last time, is that 0 factorial is defined as 1. So when we're using this formula, and we're plugging in for x equals 0, it's no problem.

So I go ahead and do that for all of the possible values of x. Of course, they're going to be functions of n in general. But for concreteness, let me choose a particular n. So we'll choose n equals 3. So we've got the probability that x is equal to 0-- is equal when n is 3 is equal to 6 over 99, and then so forth. 36 over 99 for x equals 1 and on.

And here is a picture of graphically what this probability function looks like. So here, on the horizontal axis, I have the different values that x can take on. And then I have the probabilities with which the random variable x takes on those values.

So one thing-- I want to pause for just a second and note a couple things about this graph. So by convention, when we graph probability functions, we put vertical lines beneath each point on the graph. And I suppose that just makes them easier to read.

But in probability, that's sort of a very standard convention. You'll see these sort of vertical lines instead of just dots on the graph. And then also, a piece of terminology that you should know is that each one of these points is called a point mass.

Any questions what we've done here? So this is a pretty important thing to understand. We started from this verbal description of drawing pieces of paper out of a hat that have different pizza toppings on it. And from that, we were able to write down the probability function of a random variable that we defined on that sample space. Does that make sense?

So we can do this in a more general way. So I had-- I picked the specific example of the area for menu, and that sort of let us calculate actual numerical values for our probability function. But it's also useful to recognize that this is a general type of a random variable that you might run into again. And we call it, in particular, a hypergeometric random variable.

So we say that x is a hypergeometric distribution with parameters capital N, where capital N is, in our case, the total number of toppings available for the pizza. Capital K, the total number of vegetarian toppings. And little n, the number that we're going to choose. So does this make sense to everyone? Is this clear?

And then afterwards, we have to be sure to note for which values of x there is positive probability. So for which values of x this is the formula to compute the probability. And I've been a little lazy here.

If I wanted to be absolutely correct, I would have written this as f sub x of little x is equal to this formula for these values of x, and it's equal to 0 otherwise. Sometimes we leave off the equals 0 otherwise just because we're lazy. But it's sort of understood.

So what can we say-- how do we interpret this distribution? Sometimes it is useful to think in terms of a description of an experiment or something like that when we're thinking about random variables. And the interpretation I put on the hypergeometric is that it describes the number of, quote, "successes" in n trials when you're sampling without replacement from a sample of size n whose initial probability is k over n.

So sampling without replacement. We're not putting the pieces of paper back in the hat. We pull them out, and we read them, and then that's it. And the initial sample size is n. And the initial probability of success-- in our case, a, quote, "success" is defined as a vegetarian topping. So the initial probability of success is k over n. And then, of course, that probability of success changes over the course of the experiment as you're drawing without replacement. Yep.

AUDIENCE: [INAUDIBLE] max and the min again, just [INAUDIBLE], what does that equation mean?

SARA ELLISON: OK. So it means that the-- these are the values for which x has positive probability. So it can't, obviously, have positive probability for anything less than zero. And it can't-- it also can't have positive probability if-- let me see-- n plus k minus N.

So if capital-- if the number of vegetarian toppings plus the number that you choose minus-- let me think through this. So the number you choose plus the number of vegetarian toppings minus N is positive, then that's the minimum that the x can take on.

And then the maximum it can take on is either the number of vegetarian toppings you're choosing, or-- sorry the number of toppings you're choosing, or the total number of vegetarian toppings. It can't exceed either of those. OK. Questions?

OK. Second example. So remember that we had-- in the Steph Curry example we did last time, we talked about how we could think about the number of shots he made, the number of 3-point shots that he made, as having a binomial distribution if we made certain assumptions, such as the shots were independent, and the probability of making each shot was the same for each shot.

So let's now define a random variable that's based on this experiment. And we'll call this random variable x as well. And let it be the number of 3-point shots that Steph Curry makes in the next six shots he takes.

So just like we did in the previous example, we can calculate the probability that x equals 0, x equals 1, x equals 2, and so forth up to 6. And we actually calculated a couple of those last time-- two or three of them last time-- in the example. But we can-- we have a formula here. 6 choose x times 0.44 to the x times 0.56 to the 6 minus x.

And remember where this comes from. This is the probability that he makes any particular shot. This is the probability that he misses any particular shot. This is the-- without this coefficient out front, that's the probability of any particular sequence of x shots made and 6 minus x shots missed. And then if we add this coefficient out in the front, that tells us-- that counts the number of such sequences there are. So you guys remember this from last time.

So we just use this formula. We plug into it to calculate the probabilities for all of the various values of x. And, like I said, some of these probably look familiar because we calculated them last time. I think we calculated the probability that x equals 3 last time. And maybe we calculated the probability that x equals 0 last time. I can't remember.

So graphically, what does this look like? There we go. And again, I graphed each point mass as sort of a dot and a sort of a vertical line under it.

Now, if you just glance at this, you might think it's symmetric-- this distribution is symmetric. But, in fact, it's not. I mean, if you look at the numbers, you can tell it's not. And if you look closely at my drawing, you can also tell it's not.

And, again, just like in the last example, we can generalize this example. So we can say instead of just considering events where there's a 0.44 probability of a success and a 0.56 probability of a failure, we can use more general notation and consider arbitrary probability of success as p, and also consider an arbitrary number of trials, n.

So we say that x has a binomial distribution with parameters n and p. And that's denoted using that notation. And then its probability function looks like this. And, again, I was sort of-- I was too lazy to put equals 0 otherwise, but that's understood.

OK. Does this make sense? Do you guys-- given these two examples, do you feel like you have a little bit of a sense about what a random variable is, a little bit of intuition? OK. But do feel free, as I said, to stop and ask questions if we're going through this material too quickly.

OK. So the binomial distribution comes up a lot. And it's also-- it can be very useful in lots of different settings. Here's a picture of several different probability functions from binomial distributions with different parameters.

So this picture-- in this picture, all of the parameters-- all of the p parameters are equal to 0.5. So if p is equal to 0.5, you will, in fact, get a symmetric distribution. If p is not equal to 0.5, you won't get a symmetric distribution. So here, all of these p's are equal to 0.5. And then you can see how the distribution changes as n increases.

And that one on the right is starting to look like another distribution you've probably seen before-- the normal distribution. And we'll be encountering that soon enough. And we'll also talk about the relationship between the binomial distribution and the normal distribution.

So let's step back and think a little bit more formally about this notion of a probability function. What is a probability function, and what properties does it have. Well, the probability function of x, where x is a discrete random variable, is defined as the function such that for any X, f sub x of little x is equal to the probability that X is equal to little x.

So this is just formalizing what we did in the last two examples. We just created these functions. We called them little f, called them probability functions, and denoted them little f, I guess. And all we did is we just computed probabilities that our random variable was equal to various values. And that defined our probability function.

So what are some of the properties that the probability function is going to have? Well, because it defines the probability with which random variables take on various values, it is by its nature going to have certain properties that it always satisfies.

So, in particular, the value of the probability function for any x is going to be between 0 and 1. And that's just because-- that's just by definition. It's a probability. And you can go back-- and we can go back and check the examples I did. All of the probabilities I computed were between 0 and 1.

The second property is that if you sum up over all of the possible values that the random variable can take on, then if you sum up the probability function, then that sum has to equal 1. Again, this follows directly from the fact that the probability function is sort of describing probabilities.

So let me just go back, and we can just convince ourselves. I did double-check, so it is, in fact, true. But we can convince ourselves those things do sum to 1. And the other example, those also sum to 1.

And the third thing very useful to know about a probability function is how to use it to compute probabilities. So, obviously, if you just want the probability of a particular value of x, the probability function gives it to you. You just read it off. But what if you want the probability of a whole set of different values of x?

Well, not surprisingly, you just sum up the probability function over X's in that set. And so that's going to be useful when we're given a probability function for a random variable, and we want to calculate the probability that x is in a particular set or region or something like that.

So I briefly mentioned continuous random variables a couple of minutes ago. So for continuous random variables, we're rarely going to start with a verbal description of some kind of experiment, like we did with the discrete random variables. And instead, we're typically just given a function that describes the probabilities with which a continuous random variable is in various regions. And that function is called a density.

So all of you have seen lots of examples of densities. Over on the left is a normal density. This is a triangular density. These are examples of log beta densities. So you've all seen examples of them. And we will see many additional examples and sort of learn how to manipulate them in various ways. But let me say a little bit more about this function, the density, or also known as the probability density function.

So it's an analog in a lot of ways to the discrete probability function that we were just discussing. And the next slide, we'll talk about how they're similar and different. But first, let me give you a more formal definition of what a PDF, or Probability Density Function, or density, is.

Well, so in fact, we define the continuous random variables in terms of this function. So let me just do that. A random variable x is continuous if there exists a non-negative function, f sub x, such that for any interval A in the real line, the probability that x is in A is just equal to the integral over that region A of the PDF f sub x.

So note that I've given these two functions-- the one that I said was analogous for the discrete random variable and the continuous random variable-- they have different terms. One is called a probability function. One is called a probability density function. And we do that to emphasize the fact that they're not exactly the same. They don't behave exactly the same way mathematically.

However, they're going to be-- they're analogous in a lot of ways, and they're going to be used in a lot of the same ways. And so that's why we use the same notation, this sort of f sub x of little x notation. So we use the same notation for both the discrete probability function and the continuous probability density function. Oh, and yes, that's the PDF-- that DF.

So I said I was going to go through some of the similarities and the differences between a probability function and a probability density function. So just like the probability function, the probability density function has properties induced by our earlier definition of a probability. So in particular, it's always going to be non-negative.

Now, note that it can be greater than 1. The probability function is never going to have a value greater than 1 because it is, by definition, a probability. Each point is just the probability that a random variable is equal to that point. The definition of the probability density function is a little different. So it can-- there can be regions or areas where the probability density function exceeds 1, and that's perfectly fine.

The probability function-- when you sum up the values for the probability function over all of the possible values of x you get 1. Here, the continuous analog is that we integrate over all the possible values of x, and we get 1.

And then, again, just like when I had the previous slide up with the probability function, and I said, if you want to compute the probability that x is in a particular region, you just sum up the probability function over all of the x's in that region. Well, here, you do the continuous analog, which is integrate over that region.

So we're interested in x-- the probability that x is between A and B. Oh, actually-- oh, no. I guess I have it over here. So the probability that x is in some region, capital A, which we define as an interval between A and B, all we do is we just integrate over that region.

A couple of things to keep in mind about a probability density function-- sort of relative to the probability function. So the value of a probability density function at a particular x does not have the same interpretation as a probability. In fact, the probability that the random variable x is equal to some number little x is equal to 0 for any x if x-- if big X is a continuous random variable.

So this is the sort of fundamental difference between a discrete random variable and a continuous random variable. Discrete random variable takes on particular values with positive probability-- a finite or countably infinite number of particular values with positive probability. A continuous random variable is equal to a particular value with probability zero. But instead, we talk about it being in a region, for instance, with positive probability. Yes?

AUDIENCE: I have a question where-- so if we have value of f of sub x of little x is greater than 1, then how do you integrate everything up to just 1?

SARA ELLISON: Well, because it can be greater than 1 on an interval that's shorter than 1. So like, for instance, let me draw an example. So let's take the unit interval, and we have a density that's defined on the unit interval. Well, what on the unit interval can integrate to 1?

Well, we know if we just-- let's consider this function. That integrates to 1. It never exceeds 1. But-- let's see if I can draw this in a reasonable way-- this function should also integrate to 1 over the unit interval, but it does exceed 1 for part of the time. Was there anything else I wanted to say about this? Any other questions? No? OK.

So I do want to pause for a second to say a word about terminology and notation. So I sort of did a little bit of a search on the internet and tried to find what the standard terminology and standard notation for all of these entities in probability is. And it turns out there isn't one. There is a huge range of different terms that different textbooks use and different websites use and both formal terms and informal terms.

So I just want you to be aware of that if you're looking-- if you're using a book on probability as a resource looking things up, the terminology might be different than what we're using in class. I'm going to try to be consistent-- internally consistent. But that doesn't mean that I will always be completely formal.

So, for instance, I will try to say probability function when I'm referring to-- or sorry-- yeah-- probability function when I'm referring to a discrete random variable. And I'll try to say probability density function when I'm referring to a continuous random variable. But sometimes I'll just punt and say distribution. And that's kind of an informal catch-all phrase. But do feel free to ask if you have-- if you want clarification on any of these points. Yep?

AUDIENCE: [INAUDIBLE] quick question on how to get continuous distribution [INAUDIBLE] multliply the probability of zero. So let's say you have a spike at that one point, or maybe it goes to infinity. So that uniform distribution [INAUDIBLE] integration fail or [INAUDIBLE] goes.

SARA ELLISON: Yeah. I mean, it's just not a well-behaved distribution. And so yeah. So basically, things like regularity conditions that ensure that it's intergrable might not still exist.

So what you're thinking of is like a distribution like this that's sort of getting-- that's kind of in the limit going down to a single point getting more concentrated, more concentrated.

AUDIENCE: [INAUDIBLE] break. So you have at one point just a spike.

SARA ELLISON: So what you're actually perhaps talking about is a mixed distribution. So such a-- so I haven't mentioned them, but such a thing definitely exists. So tell me if this is what you mean.

So you have a distribution for a continuous random variable. And then maybe at this one particular-- well, actually, let me do a sort an example that makes more sense. And maybe-- so maybe this is the distribution of some measurement. And then our measuring technology can't measure anything above this point. So everything-- every measurement that we would like to take at this-- above this point gets recorded at this point. Is that what you mean?

So there is a point mass here. This is like we have a continuous part of the random variable. And then this one part of the random variable is discrete. And so there is sort of positive probability at that point.

So yeah. You're absolutely right. It's just called a mixed random variable. And when we deal with mixed random variables, we just have to be cognizant of the fact that none of these-- there isn't a positive probability on any of these points, but on this one particular point there is a point mass, and we have to treat that differently. OK. Other questions? Nope.

So let me go through an example. So let a and b be real numbers, such that a is less than b. So ab defines an interval on the real line. And suppose that we have a random variable x, and it's defined in such a way that the probability of x belonging to any subinterval of S is proportional to the length of the subinterval.

So then-- and you can perhaps-- this isn't a formal proof, but you might be able to convince yourself if you think about it. The probability density function of that random variable that I've just described will take this form. So basically, it's going to have-- it's going to have a uniform density at value 1 over b minus a over the interval a to b. And it's going to be 0 otherwise.

So definitely, once you write down this PDF, and you think about this description of the random variable, you certainly can convince yourself I think that any-- the probability that the random variable X is in any subinterval of ab is going to be proportional to the length of that subinterval because this is just a constant probability. Or-- yeah, constant PDF over that interval.

So this is a special random variable. We'll see this again over the course of the semester. It is perhaps my favorite random variable. It is the uniform random variable. And it comes in handy in lots of different circumstances. And-- oh, I think actually-- I drew a picture up here on the board, but I think I may have a picture in my next slide. Yep, there's a picture.

So if you want to compute the probability of a uniform ab random variable being in some interval cd, which is a subinterval of ab. Well, we know one way that we can-- you do that because we saw that several slides ago. You can just integrate the value of the probability density function over that subinterval.

But, in fact, there's kind of an easier way to do it with a uniform distribution, a uniform random variable. We don't have to actually do the integration because this is just constant over the whole interval. So to compute the probability of any subinterval, as I said, it's just the length of the subinterval over the length of the entire support of the random variable.

So we've got probability functions, and we've got probability density functions. And for two different kinds of random variables, those describe the probabilities that the random variables are in different regions or take on different values. It's a complete description. But sometimes, it's going to be handy to express those probabilities related to a random variable in an alternative form.

Doubly handy is the fact that this alternative form has the same definition regardless of whether the random variable is discrete or continuous. And this alternative form is the cumulative distribution function. So a cumulative distribution function, which we'll often abbreviate as CDF, it's denoted with a capital F sub x. And it is defined simply as the probability that the random variable X is less than or equal to some value of little x. Same definition for both the discrete and continuous random variables.

And I should say, by the way, sometimes I often like to put the random variable as a subscript on either the probability function or the PDF or the CDF just so we-- just to keep things straight. It doesn't have to be there. So sometimes I'll omit-- there's no chance of ambiguity. Sometimes I'll omit the subscript and just use the little f for probability function or PDF and then the capital F for CDF.

So just like in the case of the probability function and the probability density function, the properties of probability are going to imply certain things about CDFs. There's certain ways that CDFs are going to behave. So, first of all, the CDF is always going to have values between 0 and 1. And, again, I mean, just-- you can look back at the definition and convince yourself that this is true.

Second thing is that the CDF is always going to be nondecreasing in X. And, again, if you think about the definition, it's just the definition that X-- the random variable X is less than or equal to little x. As little x increases, then capital F is not ever going to decrease.

The third property is that the limit as x goes to negative infinity of the CDF is equal to zero. And along with that is the limit as x goes to positive infinity of the CDF is equal to 1. So what are these CDFs going to look like?

So we know that they're always going to start at 0. So these are just values that x can take on. So the CDF is always going to start at 0. It's going to go up. It's never going to decrease. It might have flat parts-- relatively more flat parts. And then in the limit, as x goes to infinity, it's going to be equal to 1.

Now, I've drawn a CDF for a continuous random variable. This CDF doesn't have any jumps in it. A jump occurs in a CDF when there is positive probability that the random variable takes on a particular value. So for continuous random variables, the CDFs are always going to be smooth. For discrete random variables, they will always have jumps. And for a mixed random variable, they could be sort of a mixture of the two.

So what does a CDF-- maybe a typical CDF of a-- let me see. Oops. I went a little too high. So here, this might be a typical CDF for a discrete random variable. How many-- at how many points does this discrete random variable have positive probability? Three points, right? Right here, right here, and right here.

So basically, the cumulative distribution function is doing precisely what its name suggests it would do. As you go from negative infinity up to positive infinity, you're accumulating probability. So you can think of maybe the probability function kind of sitting in the background of the board. And we go along, and we get to the first point mass in the probability function. And then, at that point, the CDF jumps up the amount of the size of the point mass.

And then we go along. In this whole region, there's zero probability. So we're not accumulating any probability. We get to the second point mass in the probability function. And this is the size of the second point mass. So our CDF jumps up at that point.

So the PF/PDF, as I said, is a complete description of the probabilities associated with a particular random variable. The CDF is also a complete description of the probabilities associated with a random variable. And so, in particular, they contain the same information about its distribution just in a different form. So it stands to reason then that given the PF or the PDF one could recover the CDF. And given the CDF, one can recover the PF or the PDF.

So in the case of a discrete random variable, that can be a fairly simple process. So I just described it up here. So if I gave you this CDF, you would be able to draw the PDF-- or sorry-- the PF-- you would be able to say their point mass-- I'd have to label this, I guess, so you'd know where the point masses are.

But you'd know there was a point mass here, a point mass here, a point mass here. And you'd know how big this one and how big this one and how big this one was, and you'd be done. And likewise, if I'd given you the PF, you could have constructed this pretty easily.

For continuous random variables, these are the formulas that tell you how to move between a PDF and a CDF. And they intuitively, I think, make complete sense. They're just the continuous analogs of what I described here with the discrete random variables.

So if you want to get the CDF for a continuous random variable at a particular point, then you integrate the PDF up to that point, and that's what we have here. And if you have the CDF, and you want to get the PDF, you want to recover the PDF from it, then you take the derivative.

So does this make sense? OK, good. Oh, I should say the second-- I have to add this disclaimer. This is true provided that x is continuous and f prime exists at all but a finite number of points.

So when we were talking about probability last lecture, at some point during the lecture I said that it was going to be important for us to discuss the relationships between stochastic events. And then we talked about independence of events, and we talked about conditional probability. And then we sort of segued into Bayes' theorem and so forth.

Well, we're going to talk about analogous concepts in the context of random variables as well. But in order to do that, I first have to define a joint distribution. So in the case where we only have two random variables-- where only two random variables are involved-- we're going to call that a bivariate distribution. We'll see multivariate distributions later on in the semester.

So let me pause now to give you some indication why we care about the relationship between random variables. Well, maybe you know exactly why we do, and that's why you're in this class because that's sort of-- much of what we do in-- much of what we do in data analysis is gather repeated observations from particular joint distributions of random variables to try to figure out what their relationship is. So we do this all the time, and maybe you don't need to be convinced. But here are a couple of examples of random variables whose joint distributions we might care about.

So maybe we're interested in, for instance, the joint distribution of rainfall and crop growth in a particular field. Maybe we care about the length of the regular checkout line and the length of the express checkout line and the relationship between those two.

So, in particular, maybe we sort of know what the distribution of the length of the express checkout line is, but we want to know what the relationship between that distribution and the distribution of the length of the regular checkout line is, so we know whether to switch or something like that.

We might also be concerned about-- if we're interested in investing in a stock, we might want to know the joint distribution of the dollar/euro exchange rate and the stock price of an exporting firm that would be affected by the fluctuations in the dollar/euro exchange rate. So these are just a few examples from an infinite number of examples we could come up with for why we care about joint distributions.

So now, definition. If X and Y are continuous random variables defined on the same sample space S, then the joint probability density function of x and y-- and this is denoted little f sub xy-- is the surface such that for any region of the xy plane, the probability that xy is contained in some region A is equal to the integral of that function over the region A. So it's kind of the natural generalization of the continuous-- the definition of the continuous PDF-- natural generalization to two random variables.

So here, we're just-- we've got this function here, this small f, is a surface over the xy plane. And in order to calculate probabilities, we're going to be integrating over regions of the xy plane-- integrating that surface over regions of the xy plane.

So like before, properties of probability imply certain properties of the joint PDF, such as that it's got to integrate to 1 over the xy plane. Any individual point or one-dimensional curve has probability zero, et cetera.

So the analogous function exists for discrete random variables as well, and this is how it's defined. I'm not going to say any more about it right now. And sometimes I'm just going to give definitions in terms of continuous random variables.

So now let's consider an example where we can play around with the joint distribution. So let's suppose that after hours and hours of writing lecture notes, I develop a splitting headache. By the way, this is an entirely fictional account. And after developing that fictional headache, I rummage around in my drawer, and I find one tablet of naproxen and one of acetaminophen. So those are both things that one might take for a headache.

And, by the way, I asked my sister who went to medical school whether this was an irresponsible example for me to do in class, and she said she thought it was fine. So I take both the naproxen and the acetaminophen.

Now, let X-- random variable X-- be the effect of period of the naproxen. So I've got this headache, and I take the naproxen, and the length of time that it makes the headache go away is X-- the random variable X. And let Y be the effect of period of the acetaminophen. And suppose that this is the joint distribution for x and y.

So I want to ask the following question. What's the probability that my headache comes back within three hours? Yep.

AUDIENCE: Does that all depend on the value of lambda?

SARA ELLISON: Oh, yeah. We're going to do all the calculations conditional on lambda. Yeah. I mean, I could have put a number in there. I don't know. Maybe I should have put a number in there. But yeah. No. I decided to just do this example in a more general way. And so we'll just do-- we'll calculate all of the probabilities as a function of lambda.

So how would we go about this? Yeah?

AUDIENCE: So could you integrate that [INAUDIBLE] that's the probability with [INAUDIBLE]. So it's like the-- would it be 1 minus the integral [INAUDIBLE]?

SARA ELLISON: 1 minus the integral of--

AUDIENCE: --of [INAUDIBLE]?

SARA ELLISON: So I'm not sure exactly what you're getting at. What I'm sort of wanting you to answer is the region over which-- so you're right that we're going to integrate this joint PDF to get the probability. And what I'm driving at is what is the-- can you characterize the region over which we're going to integrate this joint PDF? Yep.

AUDIENCE: Can you use 0 to 3 and then take 1 minus the probability of that? Integrate [INAUDIBLE] of 0 to three hours?

SARA ELLISON: Integrate-- so remember, this is a surface. This is a surface over the xy plane. So when you say integrate from 0 to 3, do you mean in the x dimension or--

AUDIENCE: [INAUDIBLE] double integral.

SARA ELLISON: Yeah. A double inter-- oh, OK. Yeah. Yeah. Yeah. Yeah. Yeah. Yeah. Oh, OK. Yes, so the double integral is exactly right. Yeah. Yeah. So the-- let me see. The probability that my headache comes back within three hours is the-- yes. It's 1 minus this. Exactly.

So here, I've set up the integral so that you can see exactly what we're doing-- the probability that my headache comes back within three hours. So you're right. It should be 1 minus the probability that X is less than or equal to 3, and Y is less than or equal to 3. OK. Yep.

AUDIENCE: [? Is ?] [? it ?] correct because if X is less than 3 and Y is less than 3, that means both of the medicines are not effective within the three hours, and you will get a headache. So [INAUDIBLE] [? not ?] 1 minus [INAUDIBLE]?

SARA ELLISON: Sorry. OK. Maybe I got-- maybe I've got-- I'm confused with the negatives. So this is giving me the probability-- oh, yeah. No. This is giving me the probability that my headache will come back. Yeah. Yeah. Yeah. Yeah. Yeah. Sorry about that confusion. So the probability that my headache comes back within three hours is the integral over this region, and I have a picture of it as well.

So all we do is we take the joint PDF-- so this is the surface over the xy plane-- and we're integrating from-- we're doing the double integral over the region, the 3 by 3 box at the origin. And those dots mean that I left out the intermediate steps. But you can work on them-- work on the details at home. And, in fact, this is what you get when you do the double integral. And here is a picture. This is the region over which we're integrating.

So if I wanted to draw-- I'm not sure if I can do this in a very useful way, but let me give it a shot. So let's see. I have X over here and Y over here. And then I have f sub xy here. And this is the region over which we're integrating, and the surface looks something like this. I'm not sure how useful that is. And what we want to do is to calculate the probability just integrate this surface over this area. OK? Yep.

AUDIENCE: Sorry, can you [INAUDIBLE] why is that and not 1 [INAUDIBLE]?

SARA ELLISON: Only if you want me to get confused again and scratch my head. OK, so let's go through this. So-- OK. So X is the effective period of naproxen. So if X is equal to 3-- so let's just think about it in terms of one dimension.

So if X is equal to 3, my headache comes back in three hours. Or if X is less anything less than three, my headache will come back within three hours.

ESTHER DUFLO: [INAUDIBLE] both of them are [INAUDIBLE] both of them [INAUDIBLE], which is where [? you are ?] [INAUDIBLE]. If you had only one, then you would come back [INAUDIBLE]. If that one turns out to be-- its effectiveness cannot be less than three, but then you always have the other one. So that's why you are [INAUDIBLE]. The events have to be together [INAUDIBLE].

SARA ELLISON: And perhaps this next example will-- so this next example is sort of-- I'm changing the question a little bit, and it actually might clarify the question, or might clarify the previous question. So I'm going to change the question now and ask what if I only took the acetaminophen after the naproxen stopped working. So what question-- what-- can we put this in terms of a question about probability? What am I asking now?

AUDIENCE: [INAUDIBLE] X plus Y [INAUDIBLE]?

SARA ELLISON: That's right. It's the probability that X plus Y is less than 3 because I take the naproxen. As soon as that stops working, I take the acetaminophen. So I'm asking the question whether the sum of the two lengths is less than or equal to 3.

So now the region over which we're going to integrate this surface changes, obviously. And this is-- I've set up the integral, but I can also show you the picture of how it changes. And so now-- yep?

AUDIENCE: How do you get the probability of X plus Y is less than 3 to that integral?

SARA ELLISON: How do I set up these limits of integration? Well, so basically what I'm going to do is I'm going to draw a line in the xy plane that's defined-- that defines sort of X plus Y being equal to 3. And then I've got to set out my limits of integration. So that's this line here. And then I've got to set out my limits of integration to get that triangle.

And the way I do it-- I mean, there's actually two different ways to do it. But the way I did it was letting X go from 0 to 3. But then when you-- so as X is going from 0 to 3, then Y goes from 0 to 3 minus X, which is this line here.

So the limits of integration for Y depend on X because it's a triangle. And you can look for each value of X what the limits of integration for Y will be. Does that make sense?

So I could have done it the other-- I could have switched the order of integration, and I would have had different limits and would have gotten the same answer. Yep?

AUDIENCE: This might be a weird question, but it seems like to get these to the joint probability, you just multiply each individual probability.

SARA ELLISON: Well, you're a step ahead of us, actually. So actually-- no, you're absolutely right, and that's an astute observation. And basically that means that this joint probability is the result of two random variables that are independent. And that's how you get joint distributions of independent random variables is you multiply their individual distributions together.

So we didn't-- you didn't have to know that for this example. I just gave you the joint distribution. But if you knew about independence of random variables, and you knew how to get joint distributions when they were independent, then I could have given you the individual distributions, and then you could have computed what the joint distribution was and then done the problem like that. Yeah.

So let's see. So one final and important point I want to make about this example is that this example sort of suggests something very powerful that we can do with random variables. So let's suppose I want to define a new random variable, and I call my random variable z. And that's the total effect of life of the naproxen and the acetaminophen taken sequentially.

So, I mean, it's certainly a reasonable random variable to define. How could I start to think about how that random variable is distributed? Well, in fact, we've already figured out exactly how it's distributed. So were you going to say-- oh, OK. So we've actually already figured out how it's distributed.

So the CDF of this new random variable, by definition, is just equal to the probability that Z is less than little z. It's just the definition of a CDF. That is equal to-- by definition of our new random variable, that's equal to the probability that X plus Y is less than or equal to little z, and that is equal to that quantity, which we just computed for z greater than 0.

And so once we've got the CDF of this new random variable z, then we know that we can also get the PDF of the new random variable z. And how do we do that? Well, we just take the derivative. So the PDF of this new random variable is just equal to lambda squared z times e to the negative z lambda.

So the reason why this is a very important example-- and the concepts embodied in it are going to come up many times over the course of the semester. The reason why it's so important is that when we start studying statistics, statistics are-- a statistic is nothing other than functions of random variables. That's what a statistic is.

And so if we want to understand how statistics behave, we have to understand how functions of random variables behave. And this is the first step we're going to take towards this. This is just one example.

But this lets us know that if we have two random variables, and we-- x and y-- and we know how they're distributed, we have some hope of figuring out how some arbitrary function of them is also distributed. And, in fact, we're typically going to be able to figure out how the sum of random variables is distributed. And that's going to be very useful for us.

And so that's one reason why I like this example a lot is to make the point that we can, with the tools we have already, start thinking about how to deal with functions of random variables. OK.